Statistical Inference and Computational Physics

AUTHOR(S):
伊庭, 幸人

CITATION:
伊庭, 幸人. Statistical Inference and Computational Physics. 物性研究 2000, 73(5): 872-875

ISSUE DATE:
2000-02-20

URL:
http://hdl.handle.net/2433/96780

Statistical Inference and Computational Physics

統計数理研究所　伊庭幸人 [1]

統計的情報処理と計算物理の境界領域についての筆者の研究 (1987-1999) をサーベイした．まずはじめに，初期に物理学会等で発表した研究について話し，また有限温度推定との出会いについて述べた．次に現在（あるいは少し過去）の話題として，与えられた形状のヘテロポリマーを設計する問題を紹介した．最後に，現在（あるいは未来）の課題として，「逆方向」の問題，すなわち，統計科学の思想や方法を計算物理・大規模数値計算へ応用する可能性について議論した．

1 過去

筆者がこの分野に興味を持ったのは，1987年ごろ，Geman たちの論文を読み，また統計数理研究所での情報量統計学やベイズモデルの研究に触れてからである．当時，Simulated Annealing に興味があったが，有限温度の統計力学との概念的なつながりの弱さ（確率分布が単に計算の便宜にのみ使われていること）に不満を感じていた筆者にとって，ベイズ統計の枠組みは魅力的であった．また，統数研の人々の仕事（たとえばAICによる分割表の自動解析）は人工知能への新しい展望を感じさせた．この周辺についての手探りでの研究は，1987年から1990年代のはじめにかけて，何回か物理学会などで発表した（下の表）．これらの講演では,「有限温度での情報処理」に相当する内容にも触れているが，表の時期には，物理学会のニューロ関係の人の反応はあまりなかった（もっとも，これらの多くを話したのはニューロのセッションではなく，物基統計の「その他」セッションなどであったように思う.）．

物理学会での口頭発表 (1987-1989)	
ベイズ推定と統計物理	1987 秋
統計物理とベイズ推定	1988 Apr.
モンテカルロ法と統計学: 分類の問題を例として	1989 Mar.
情報理論とレプリカ法	1989 Oct.
画像処理における平均場法	1989 Oct.
分布と最適解：anneal 無用論をめぐって	1989 Oct.

その後，1993-1995 に「ベイズ統計と統計物理」,「ベイズ統計の立場からみた Nishimori line [2]」などのタイトルで4件発表.

[1] E-mail: iba@ism.ac.jp

[2] この話題については，Y.Iba, J.Phys.A **32**(1999) 3875-3888 参照.

物理関係の研究会での発表 (1988-1994)	
Bayesian Statistics and Statistical Mechanics	YKIS88' (1988)
ベイズ統計と統計物理	複雑系 2(基研) (1993)
Prediction, Complexity and Demons	複雑系 3(基研) (1994)

当時の研究の一部は以下の論文で発表したが，筆者が根性なしであったため，急速に発展したこの分野に大きく寄与することはできず，残念であった．

■ Iba,Y. 1989 (YKIS88 の proceedings)
Bayesian statistics and statistical mechanics, *Cooperative Dynamics in Complex Physical Systems*, H. Takayama, ed., Springer-Verlag, Berlin, 1989, 235–236.
■ 伊庭幸人 1991
メトロポリス的モンテカルロ法の巨視的パラメータ推定への応用 －２次元イジング模型の場合－, 統計数理, **39-1** (1991), 1–21.
■ 伊庭幸人 1991
メトロポリス的モンテカルロ法の擬ベイズ法への応用 －変化点問題を例として－, 統計数理, **39-2** (1991), 225–244.
■ Iba,Y. 1999
Bayesian Classification with Relational Data,
ISM Research Memo No.440 (1992) (unpublished)
→ ISM Research Memo No.731 (1999) (to be submitted).

ここ数年，筆者はこの方面の研究から遠ざかる方向にある．統計物理のアルゴリズムの統計科学への導入という面では，一番エキサイティングな部分は終わったのではないかという気がするし，人工知能研究の流れも９０年代からは別の方向が伸びてきているので，以前ほどの興奮はなくなってきた．というよりも，同じようなセンスを持つ競争相手（物理学者に限らない）が増えすぎて，中々太刀打ちできなくなってきている．いわゆる理論の人，解析的な計算をしたり相図を描いたりする人にとってはこれからが稼ぎ時なのかもしれないが，思い付きや企画力が売り物の筆者にとってはつらいところである．

そうはいっても，画像問題での基本的な疑問（ハイパーパラメータ推定に関する問題，事前分布に相転移がある場合の解釈の問題など）はどうもまだすっきり解決できていないし，いくつか試みてみたいアイディアもあるので，もし何か面白い結果を出すことができたら，近い将来，この分野に復帰したいと考えている．

2 現在（少し過去）

この部分では,「与えられた形状に折りたたまれるポリマーを設計する」というパズル (inverse folding, protein design) について話した．モノマーの種類とその間の相互作用は与えられているとして，モノマーの1次元配列を求めるというのが代表的な問題である．これは,「与えられた配列のポリマーがどういう形状に折りたたまれるか」という物理の問題 (folding) の「逆問題」になっている[3]．本当のたんぱく質についてこれができれば，実用的にも生物の理解という点でも重要であるが，とりあえず，われわれが考えるのはおもちゃの問題（たとえば格子上の0－1配列のポリマーで“1”どうしは引き合うとする）である．Kurosky と Deutsch(1995,1996) はこの問題を，形状 $\tilde{r}$ を与えて，熱平衡分布

$$p(r|\sigma) = \frac{\exp(-\beta E(r|\sigma))}{Z(\sigma)}$$

が $r = \tilde{r}$ のところに山を持つような配列 σ を探す問題として定式化した．これは，統計学でいう「最尤推定」あるいは同じことであるが「ボルツマンマシン学習」の問題と同じ形をしている．このことはすでに Deutsch たちが指摘しているが，われわれはモノマーの「値」に形式的に連続値を許すことで，ボルツマンマシンの学習方程式に相当する式（デザイン方程式）を定義し，それを用いて上の式 (但し $1/\beta \sim 0$ とする) を最適化するというアプローチを考案した．詳細及び文献は以下を参照されたい．

Iba, Y., Tokita, K. and Kikuchi, M. 1998
Design equation : A novel Approach to Heteropolymer Design,
J.Phys.Soc.Japan, **67**, 3985-3990 (1998)

なお，当日は上記論文の共著者の時田氏も研究会に参加され，活発な討論が行われた．

3 将来

「過去」のところで触れた研究の問題意識は「統計物理の手法を（広義の）統計学の問題に応用する」というものであった．ここでは逆に「統計科学の考え方を大規模数値計算に役立てる」ことを考えてみる．標語的にいえば「Computation as Inference Process」ということになる[4]．

例として，モンテカルロ法を考えてみよう．従来の感覚では「統計学の応用」は誤差の推定とか，不偏推定量の計算とか，重要ではあっても周辺的なものに留まる．しかし，最近では，より本質的な部分に統計的な考えが利用されるようになってきている[5]．たとえば,

[3] 一般に，物理の問題の「逆」は，物理の世界と情報の世界の接点にあると考えられる点で興味深い．

[4] このような思想は田辺国士氏（統数研），また，川人光男氏（ATR）のそれに学んだところが大きい．

[5] 以下で議論する方向とはちょっと違うが,「計算結果をどのように解析・表現するか」という問題も重要である．この方向については，基研研究会「モンテカルロ法の新展開」(1999.2) の報告（「物性研究」に近く掲載予定）を参照されたい．

量子モンテカルロ法の計算結果からダイナミカルな量を解析接続によって得る際に，最大エントロピー法などの統計的手法が有効に使われているのはその例である．ここで，欲しい量と間接的に関係した量を測定して，それから欲しい量を復元するのは，逆問題のひとつの典型である（わかりやすい例としてはCTスキャンがある — いろいろな方向への射影を撮影してそれから人体の横断面を再構成する）．マルチカノニカル・モンテカルロ法では，本来サンプルすべき分布とは違う分布からサンプリングした結果から，欲しい量を「reweighting」によって再構成する．マルチカノニカル法では，従来のヒストグラム法と違って，サンプルの仕方（アンサンブル）自体を学習によって逐次的に構成するが，これは一種のActive Learningと解釈できる．

統計科学の考え方は，高次代数方程式，強制振動子法，負符号問題などさまざまな方面で応用可能なのではないかと思う．その中で現代的な統計学の考え方，情報量規準や階層ベイズモデルの手法をどう生かしていくかは興味深い問題である．

□□「過去」（および「将来」）については，筆者による以下の解説も参照されたい□□

◇ 文献案内の形で，筆者が考えたことや学んだことを解説したもの：
統計科学 文献案内
「物性研究」**71-6** (1999年3月号) 923-945.

◇ ベイズモデル，特に平滑化や画像再構成についての解説と議論：
学習と階層 — ベイズ統計の立場から —
「物性研究」**65-5** (1996年2月号) 657-677.

◇ ベイズ統計と有限温度推定，交換モンテカルロ法についてのメモ：
ベイズ統計と統計物理 -有限温度での情報処理-
「物性研究」**60-6** (1993年9月号) 677-699.
追加と訂正：「物性研究」**65-5** (1996年2月号) 678-685.

◇ 少し変わった統計学（主にモデル選択理論）の解説の試み：
モデル選択とその周辺
「物性研究」**72-1** (1999年4月号) 1-20.

◇ 順列組み合わせの思想とその可能性について：
無時間の思想
「物性研究」**71-2** (1998年11月号) 189-208.

◇ 統計学者向けのモンテカルロ法の解説（オリジナルの例も含む）：
マルコフ連鎖モンテカルロ法とその統計学への応用
統計数理 **44** No.1 49-84 (1996).